SPIM / INSERM ERM 0202
December 6, 2004

The Unified Medical Language System

A two-level structure

NATIONAL
LIBRARY OF
MEDICINE

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

Outline

- Background
  The Unified Medical Language System

Two themes:
- Assessing consistency between SN and Meta
- Specifying Meta relationships from SN relationships

- Three studies
  - Metathesaurus vs. Semantic Network relations in the domain of cardiology
  - Semantics of co-occurrence relations
  - Consistency of hierarchical relations between Metathesaurus and Semantic Network

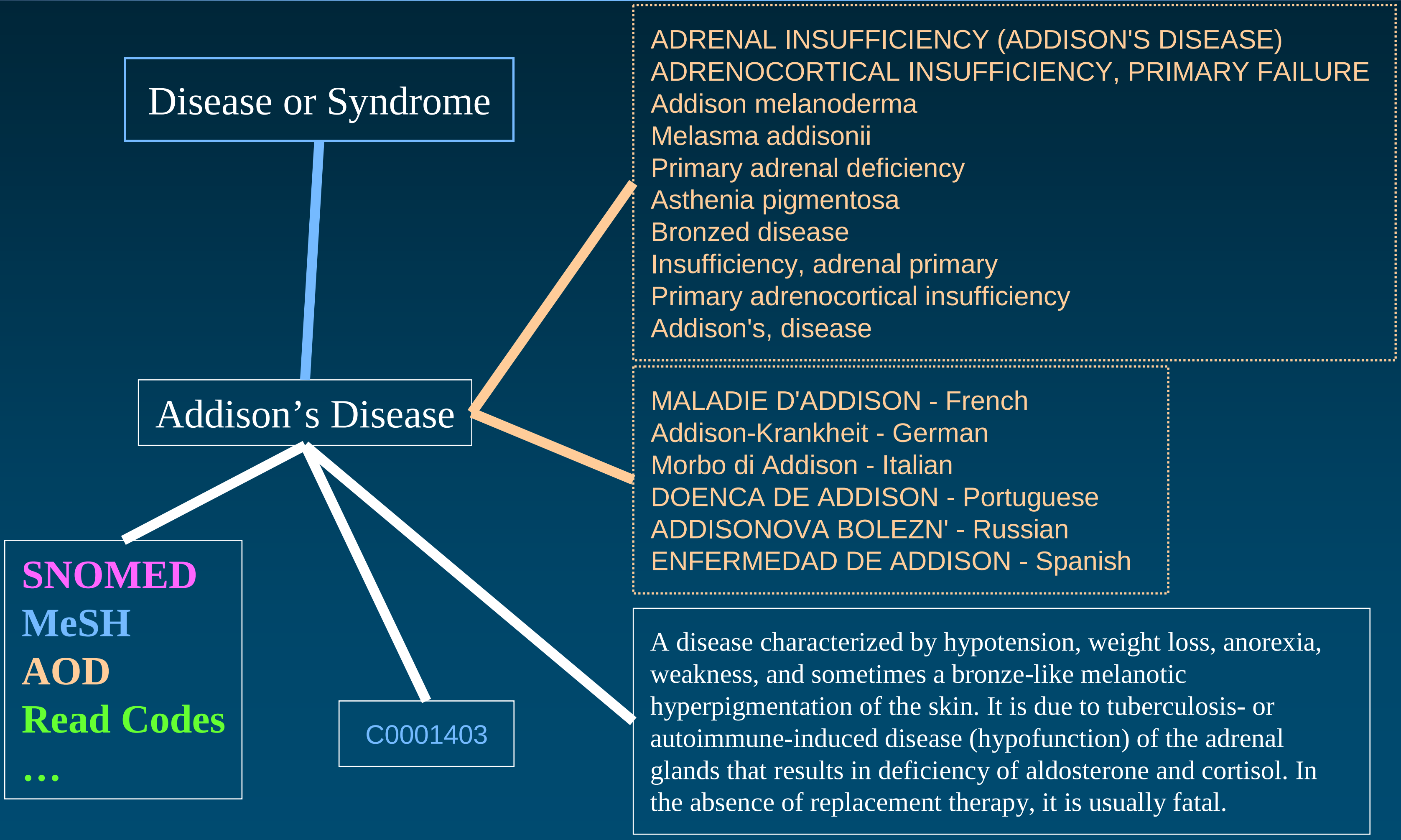
Background

The Unified Medical Language System

UMLS: 3 components

- Metathesaurus
  - Concepts
  - Inter-concept relationships
- Semantic Network
  - Semantic types
  - Semantic network relationships
- Lexical resources
  - SPECIALIST Lexicon
  - Lexical tools

Addison’s Disease: Concept



Disease or Syndrome

Addison’s Disease

ADRENAL INSUFFICIENCY (ADDISON'S DISEASE)
ADRENOCORTICAL INSUFFICIENCY, PRIMARY FAILURE
Addison melanoderma
Melasma addisonii
Primary adrenal deficiency
Asthenia pigmentosa
Bronzed disease
Insufficiency, adrenal primary
Primary adrenocortical insufficiency
Addison's, disease

MALADIE D'ADDISON - French
Addison-Krankheit - German
Morbo di Addison - Italian
DOENCA DE ADDISON - Portuguese
ADDISONOVA BOLEZN' - Russian
ENFERMEDAD DE ADDISON - Spanish

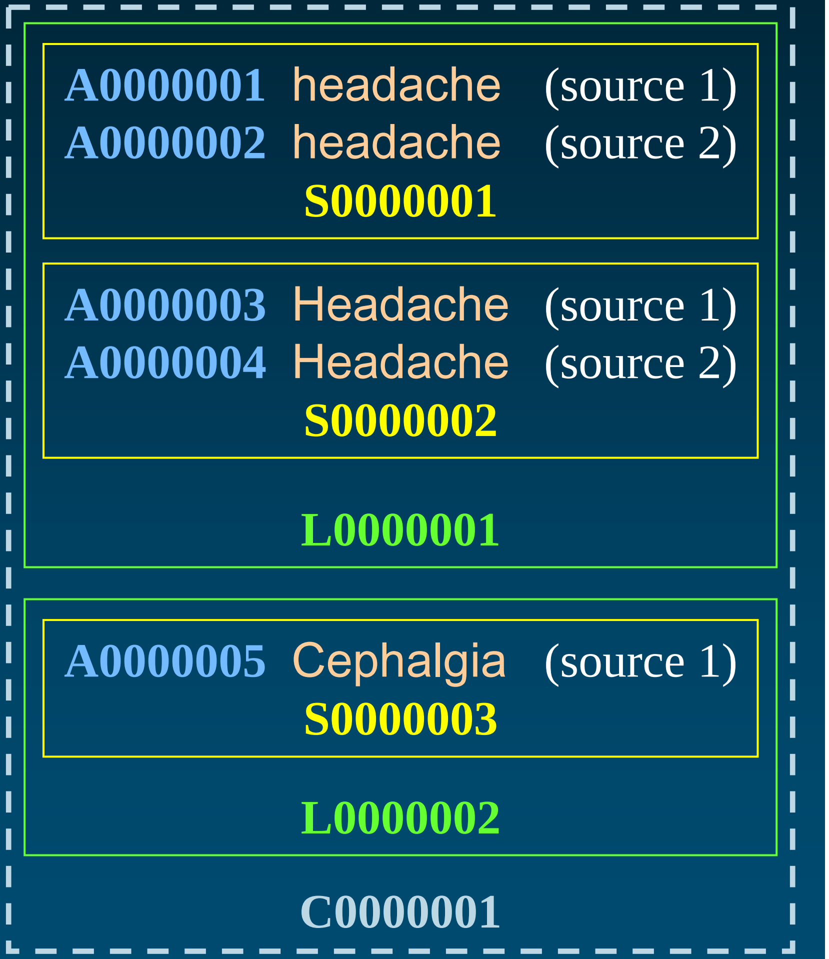
SNOMED
MeSH
AOD
Read Codes
…

C0001403

A disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin. It is due to tuberculosis- or autoimmune-induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol. In the absence of replacement therapy, it is usually fatal.

Metathesaurus Concepts (2004AB)

- Concept (> 1M) CUI
  - Set of synonymous concept names
- Term (> 3.8 M) LUI
  - Set of normalized names
- String (> 4.3M) SUI
  - Distinct concept name
- Atom (> 5.1M) AUI
  - Concept name in a given source



A0000001 headache (source 1)
A0000002 headache (source 2)
S0000001

A0000003 Headache (source 1)
A0000004 Headache (source 2)
S0000002

L0000001

A0000005 Cephalgia (source 1)
S0000003

L0000002

C0000001

Metathesaurus Relationships

- Symbolic relations: ~9 M pairs of concepts
- Statistical relations : ~7 M pairs of concepts (co-occurring concepts)
- Mapping relations: 100,000 pairs of concepts

- Categorization: Relationships between concepts and semantic types from the Semantic Network

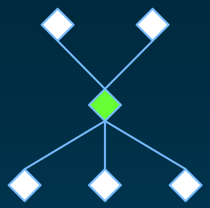
Symbolic relations

- Relation
  - Pair of “atom” identifiers
  - Type
  - Attribute (if any)
  - List of sources (for type and attribute)
- Semantics of the relationship:
  defined by its type [and attribute]

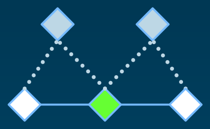
Source transparency: the information
is recorded at the “atom” level

Symbolic relationships Type

- Hierarchical
  - Parent / Child PAR/CHD
  - Broader / Narrower than RB/RN



- Derived from hierarchies
  - Siblings (children of parents) SIB



- Associative
  - Other RO

- Various flavors of near-synonymy
  - Similar RL
  - Source asserted synonymy SY
  - Possible synonymy RQ

Symbolic relationships Attribute

- Hierarchical
  - isa (is-a-kind-of)
  - part-of
- Associative
  - location-of
  - caused-by
  - treats
  - …
- Cross-references (mapping)

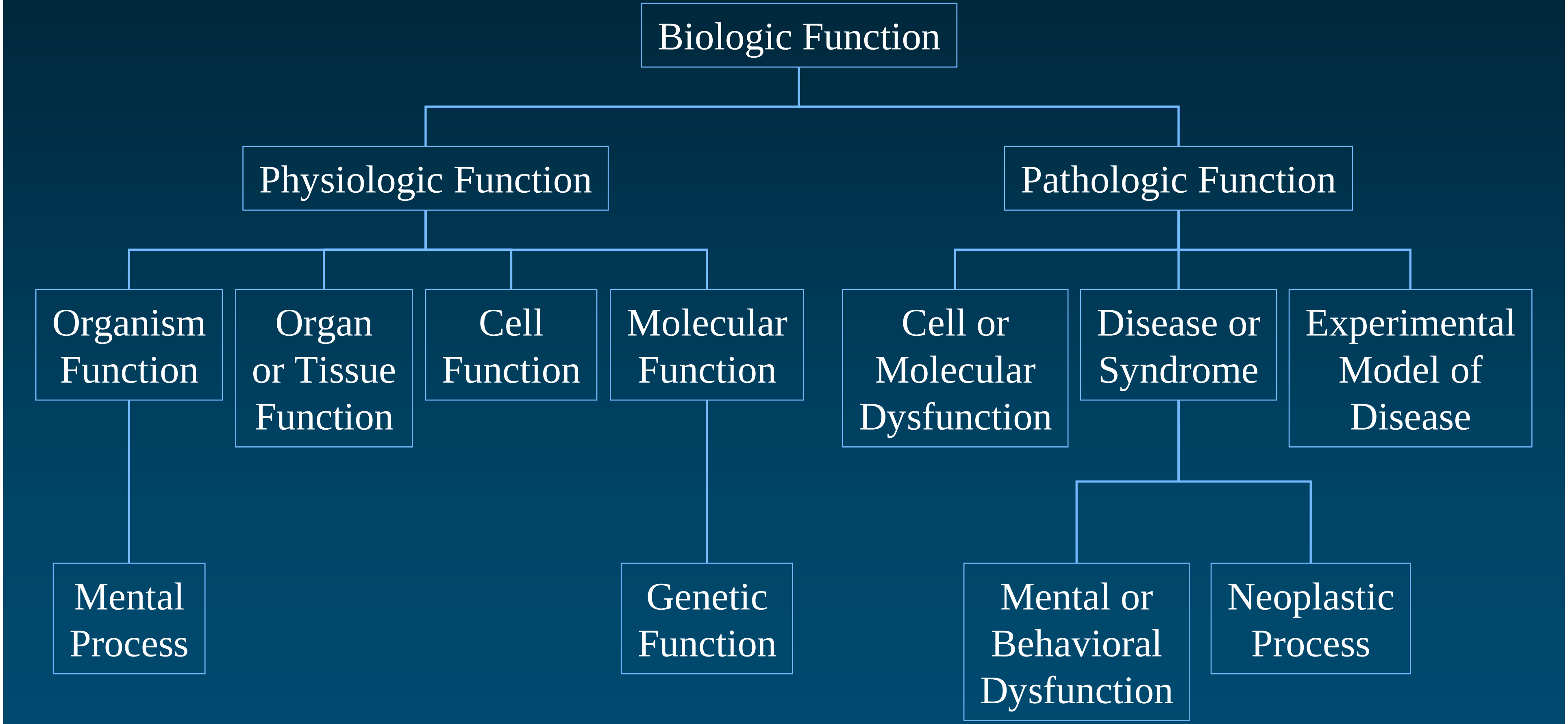
Semantic Network

- Semantic types (135)
  - tree structure
  - 2 major hierarchies
    - Entity
      - Physical Object
      - Conceptual Entity
    - Event
      - Activity
      - Phenomenon or Process

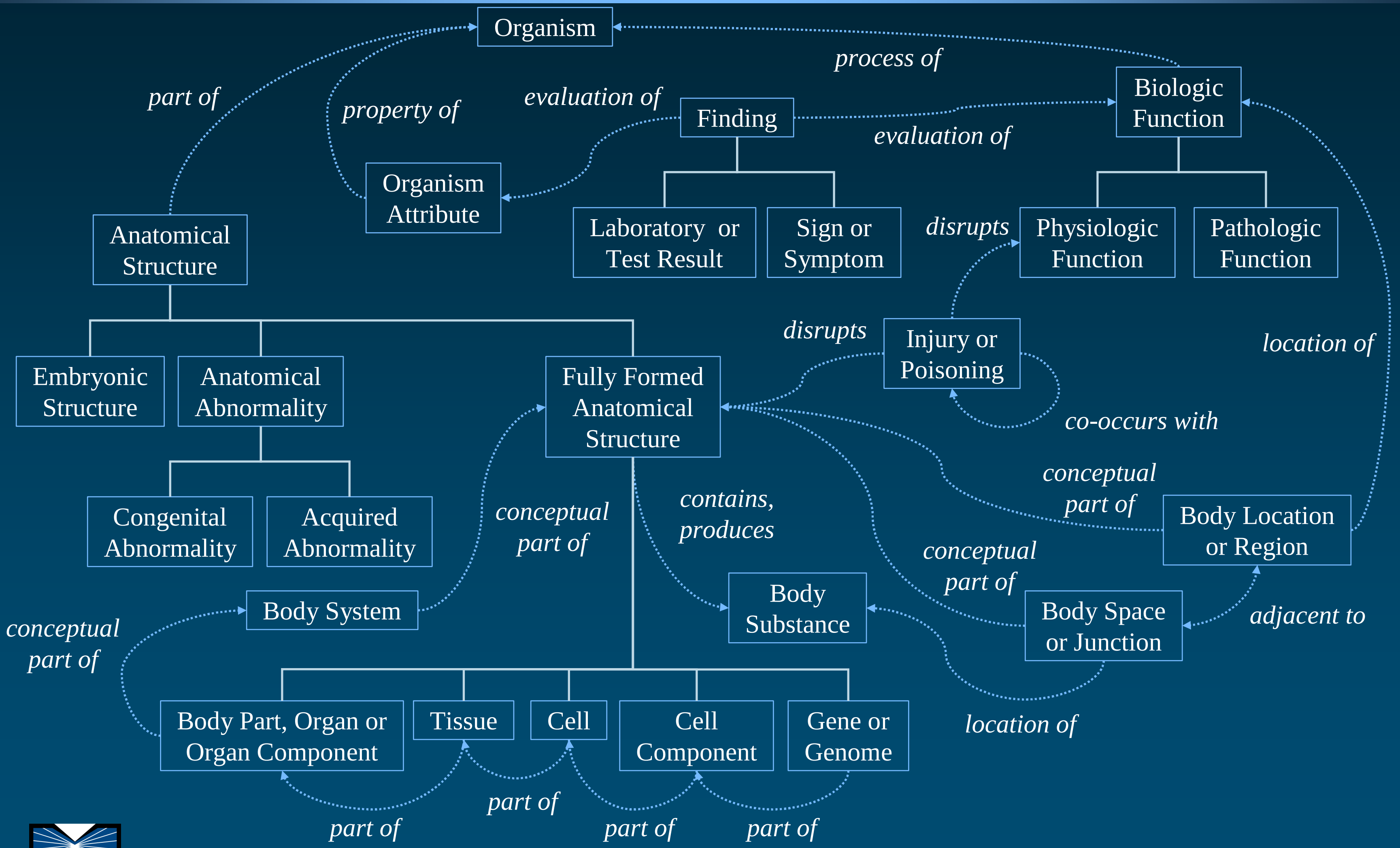
Semantic Network

- Semantic network relationships (54)
  - hierarchical (isa = is a kind of)
    - among types
      - Animal *isa* Organism
      - Enzyme *isa* Biologically Active Substance
    - among relations
      - treats *isa* affects
  - non-hierarchical
    - Sign or Symptom *diagnoses* Pathologic Function
    - Pharmacologic Substance *treats* Pathologic Function

“Biologic Function” hierarchy (isa)



Associative (non-isa) relationships

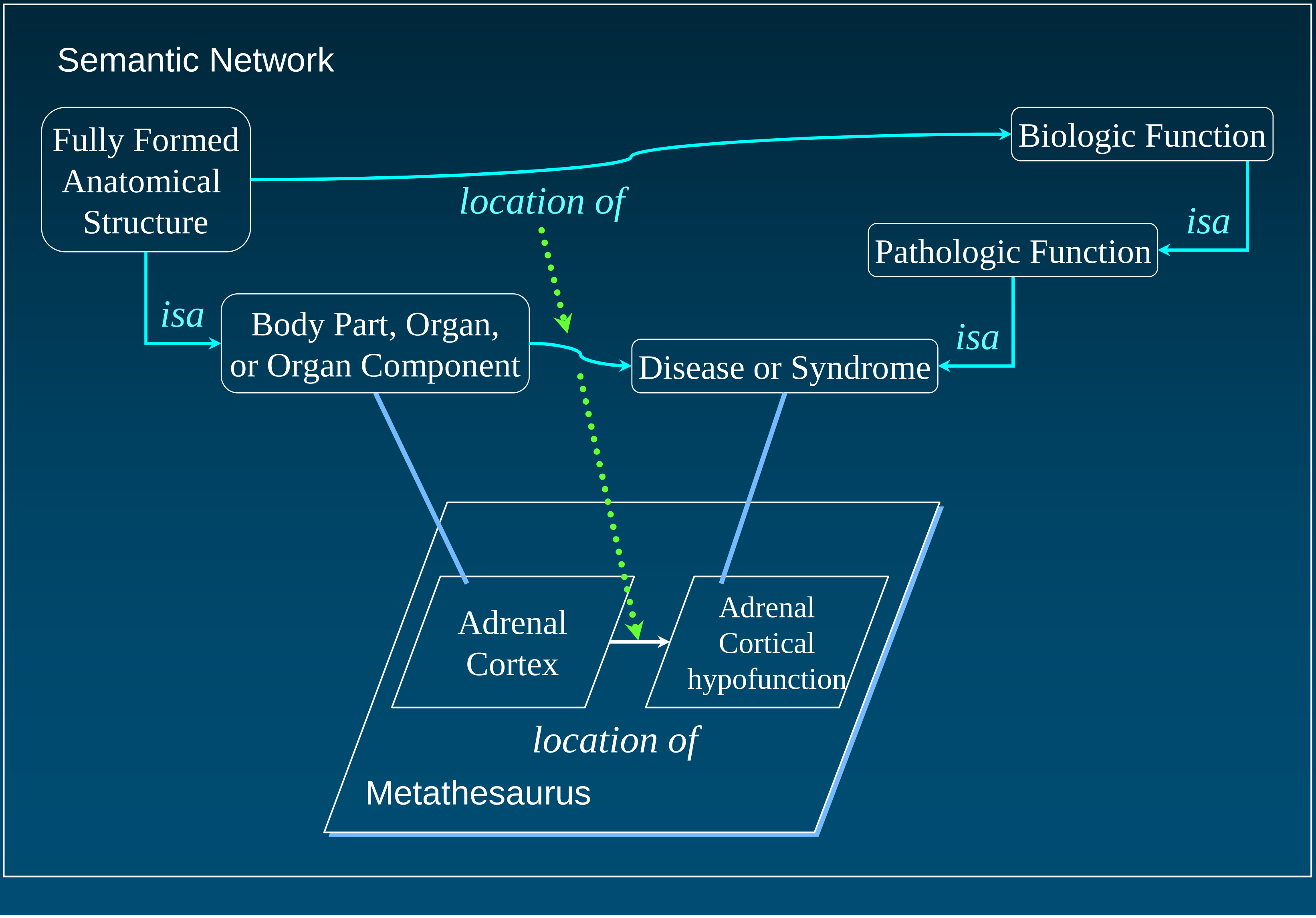


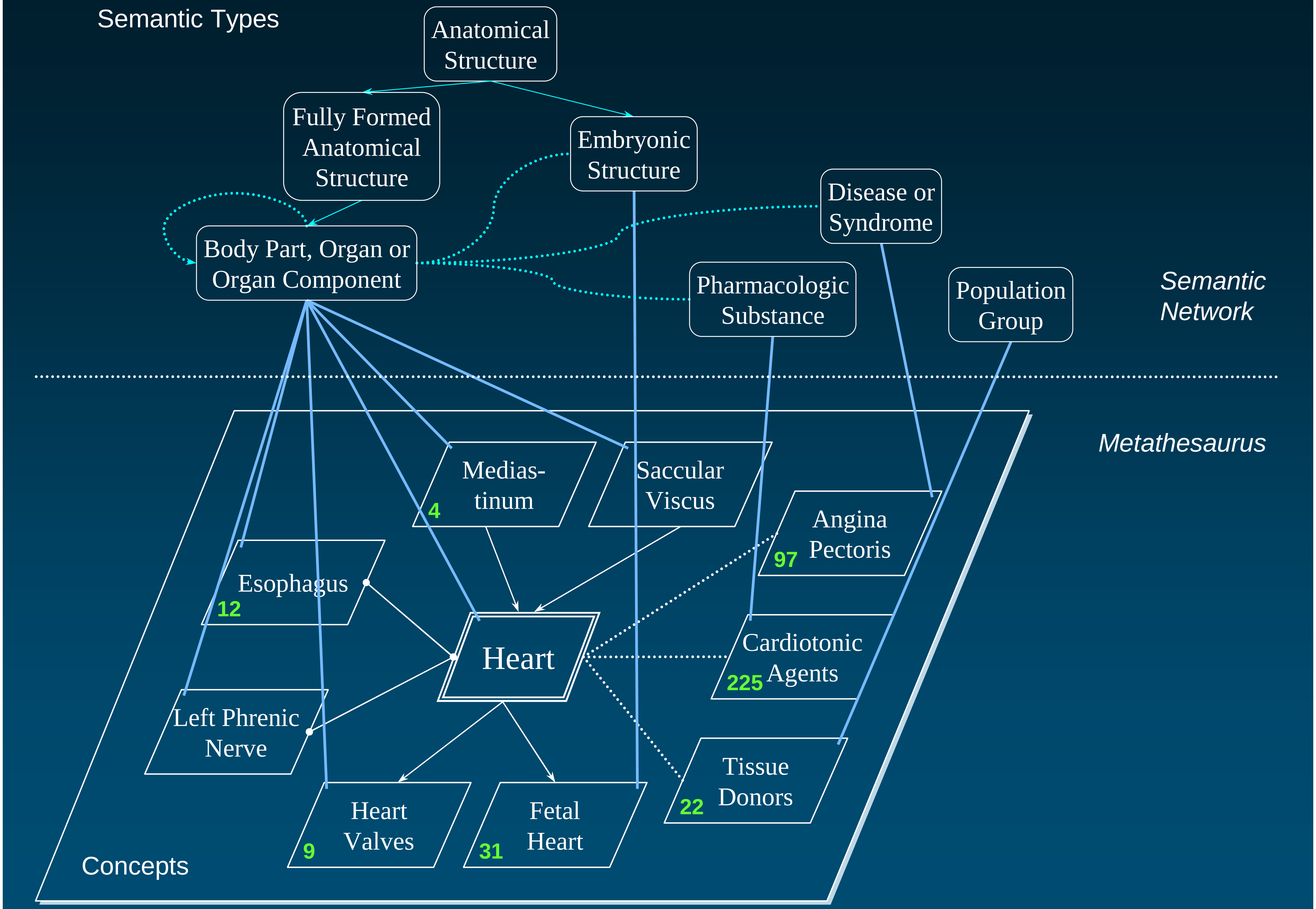
Why a semantic network?

- Semantic Types serve as high level categories assigned to Metathesaurus concepts, *independently of their position in a hierarchy*

- A relationship between 2 Semantic Types (ST) is a possible link between 2 concepts that have been assigned to those STs
  - The relationship may or may not hold at the concept level
  - Other relationships may apply at the concept level

Relationships can inherit semantics





Semantic Types

Anatomical Structure

Fully Formed Anatomical Structure

Embryonic Structure

Body Part, Organ or Organ Component

Disease or Syndrome

Pharmacologic Substance

Population Group

Semantic Network

Metathesaurus

Medias- tinum

Saccular Viscus

Angina Pectoris

Esophagus

Heart

Cardiotonic Agents

Left Phrenic Nerve

Tissue Donors

Heart Valves

Fetal Heart

Concepts

UMLS links Summary

- Semantic network relationships
  - Hierarchical or associative
  - General (definitional) knowledge
  - May or may not hold at the concept level
- Categorization
  - Links each concept to (at least) one broad category
  - Either *isa* or *is an instance of* relationships
- Interconcept relationships
  - Hierarchical, associative or statistical
  - Factual knowledge

Motivation

- Metathesaurus relations are expected to be consistent with the corresponding relations in the Semantic Network
- Many Metathesaurus relations
  - are underspecified (no RELA)
  - have no semantics (co-occurrences)

and could be refined with the Semantic Network

Three studies

- Metathesaurus vs. Semantic Network relations in the domain of cardiology (consistency and refinement)
- Semantics of co-occurrence relations
- Consistency of hierarchical relations between Metathesaurus and Semantic Network

Metathesaurus vs. Semantic Network relations in the domain of cardiology

McCray A.T, Bodenreider O.
A conceptual framework for the biomedical domain.
In: Green R, Bean CA, Myaeng SH, editors. *The semantics of relationships: an interdisciplinary perspective*.
Boston: Kluwer Academic Publishers; 2002. p. 181-198.

Motivation

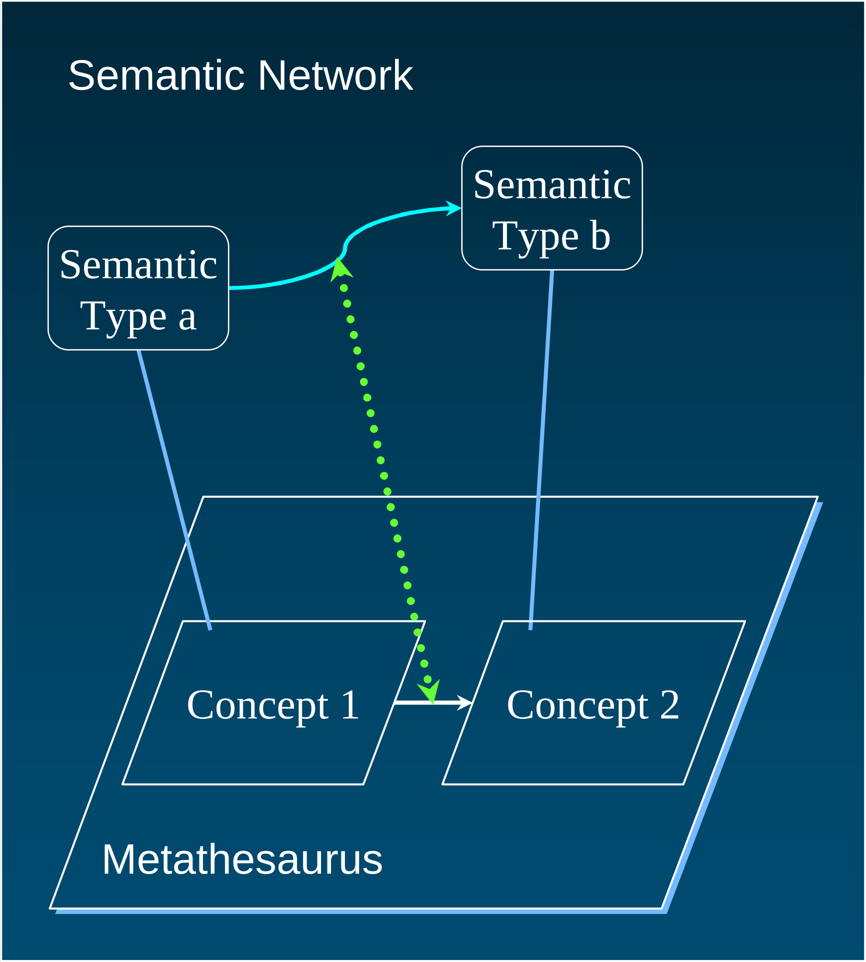
- Check the consistency of the two levels
  - Semantic network
  - Metathesaurus
- Check the consistency between
  - Semantic network relationships
  - Interconcept relationships
- Discrepancies may indicate
  - Inaccurate relationship
  - Inaccurate categorization

Motivation

- More generally
  - The Semantic Network represents some kind of upper-level ontology of the biomedical domain
  - The organization of Metathesaurus concepts
    - is *expected* to be compatible with the upper level
    - is *required* to be compatible with the upper level if reasoning is to be supported

Methods

- For each pair of related concepts
  - Get their semantic types
  - Get all the "expanded" semantic network relationships between the two semantic types (transitive closure)
  - Compare
    - Interconcept relationship
    - Sem. Net. relationships



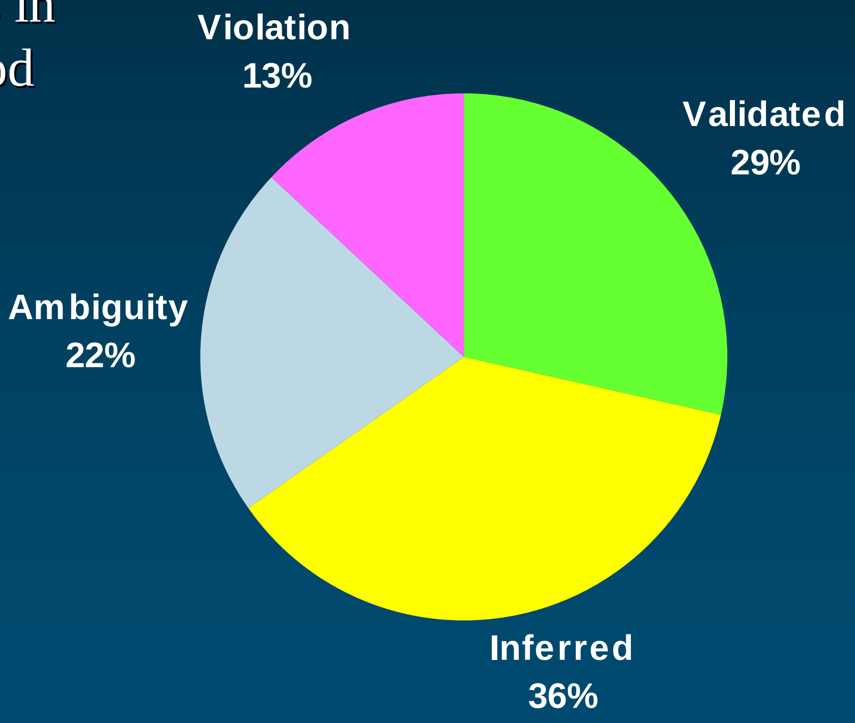
Methods

- Possible outcome
  - ICR = SNR → validate
  - ICR descendant of SNR → validate
  - ICR and SNR not compatible → reject
  - Unspecified ICR (no RELA) → infer/reject
  - ICR not in the Semantic Network

ICR: Inter-concept relationship
SNR: Semantic Network relationship

Results

- 6894 interconcept relationships
  - among the 3764 concepts in the semantic neighborhood of "Heart"



Discussion

- Interconcept relationships recorded in the Metathesaurus are not censored
- The Semantic Network
  - Provides semantic constraints
  - Can be used to select Metathesaurus relationships that are “semantically sound”
- Limitations
  - Ambiguous SN relationships
  - Unspecified Metathesaurus relationships
  - Need for some degree of manual review

Semantics of co-occurrence relations

Burgun A, Bodenreider O.
Methods for exploring the semantics of the relationships between co-occurring UMLS concepts.
Medinfo; 2001. p. 171-175.

Co-occurrence Overview

- Co-occurrence between MeSH descriptors in MEDLINE citations
- 7 M pairs of co-occurring concepts
- Implicit semantics
- The UMLS provides knowledge for helping make this relationship explicit
  - Corresponding symbolic knowledge (Metathesaurus)
  - Categorization (Semantic Network)

An example from MEDLINE

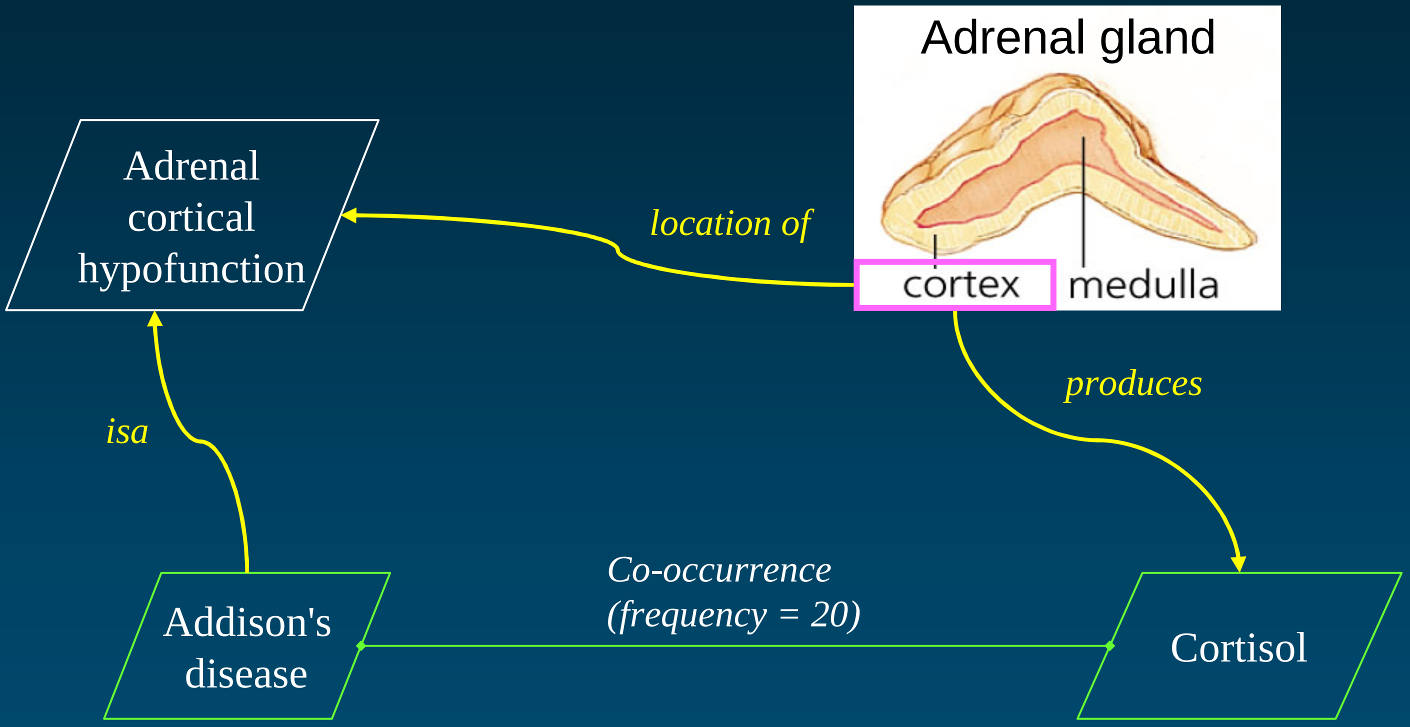
Cugini P, Letizia C, Cerci S, Di Palma L, Battisti P, Coppola A, Scavo D.
A chronobiological approach to circulating levels of renin, angiotensin-converting enzyme, aldosterone, ACTH, and cortisol in Addison's disease.
***Chronobiol Int* 1993 Apr;10(2):119-22**

This study deals with a chronobiological approach to the circadian rhythm of the renin-angiotensin-aldosterone system (RAAS) and the ACTH-cortisol axis (ACA) in patients with Addison's disease (PAD). The aim is to explore the mechanism(s) for which the circadian rhythmicity of the RAAS and ACA takes place. The study has shown that both the RAAS and ACA are devoid of a circadian rhythm in PAD. The lack of rhythmicity for renin and ACTH provides indirect evidence that their rhythmic secretion is in some way related to the circadian oscillation of aldosterone and cortisol. This implies a new concept: a positive feedback may be included among the mechanisms which chronoregulate the RAAS and ACA.

PMID: 8388783, UI: 93272348

- Addison's Disease/physiopathology
- Addison's Disease/blood*
- Adolescence
- Adult
- Aldosterone/blood*
- Circadian Rhythm*
- Corticotropin/blood*
- Female
- Human
- Hydrocortisone/blood*
- Male
- Middle Age
- Peptidyl-Dipeptidase A/blood*
- Renin/blood*

Example



Adrenal gland

cortex

medulla

Adrenal cortical hypofunction

location of

isa

produces

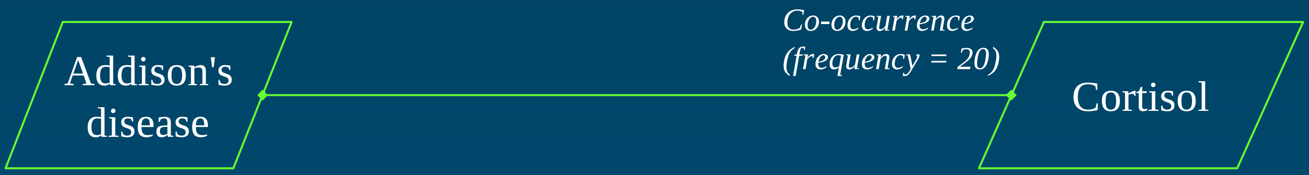
Addison's disease

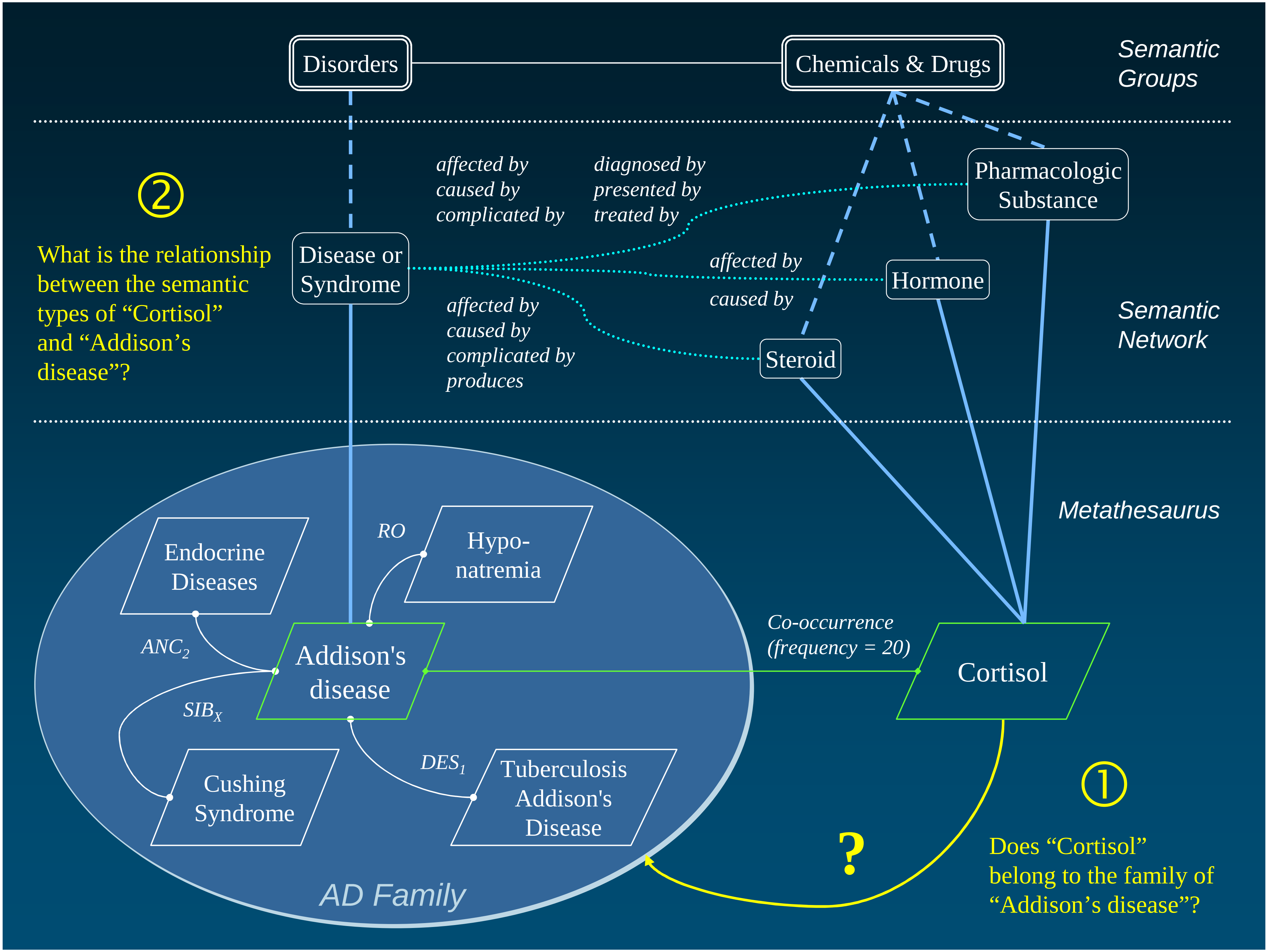
Co-occurrence (frequency = 20)

Cortisol

Methods

- Based on Metathesaurus relationships
  - Does “Cortisol” belong to the family of “Addison’s disease”?
- Based on Semantic Network relationships
  - What is the relationship between the semantic types of “Cortisol” and “Addison’s disease”?





② What is the relationship between the semantic types of "Cortisol" and "Addison's disease"?

① Does "Cortisol" belong to the family of "Addison's disease"?

Results

- Family
  - Only 6% of the relationships between co-occurring concepts correspond to symbolic relationships recorded in the Metathesaurus
- Semantic groups
  - The semantics of the relationship often remains ambiguous
  - Most frequent association:
"Chemical & Drugs" to itself

Consistency of hierarchical relations between Metathesaurus and Semantic Network

Bodenreider O, Burgun A.
Aligning knowledge sources in the UMLS: Methods, quantitative results, and applications.
Medinfo; 2004. p. 327-331.

Concepts vs. semantic types

- Semantic types
  - 135
  - High-level categories
    - *Cell*
    - *Injury or Poisoning*

- Concepts
  - 1 M
  - Mostly fine-grained
    - *Postganglionic neuron*
    - *Closed fracture of shaft of femur*
  - But not all
    - *Cells*
    - *Injuries*
    - *Poisoning*

Objective

Investigate the equivalence between

- Semantic types
- Concepts

Approaches

- Aligning knowledge structures
- Conventional approaches ❶
  - Compare names → Lexical similarity
  - Compare definitions
  - Compare relations
- Specific to UMLS
  - Categorization relation between concepts and semantic types
  - Hierarchical structure among concepts ❷
  - Compare sets of concepts → Conceptual similarity

➊ Lexical similarity Method

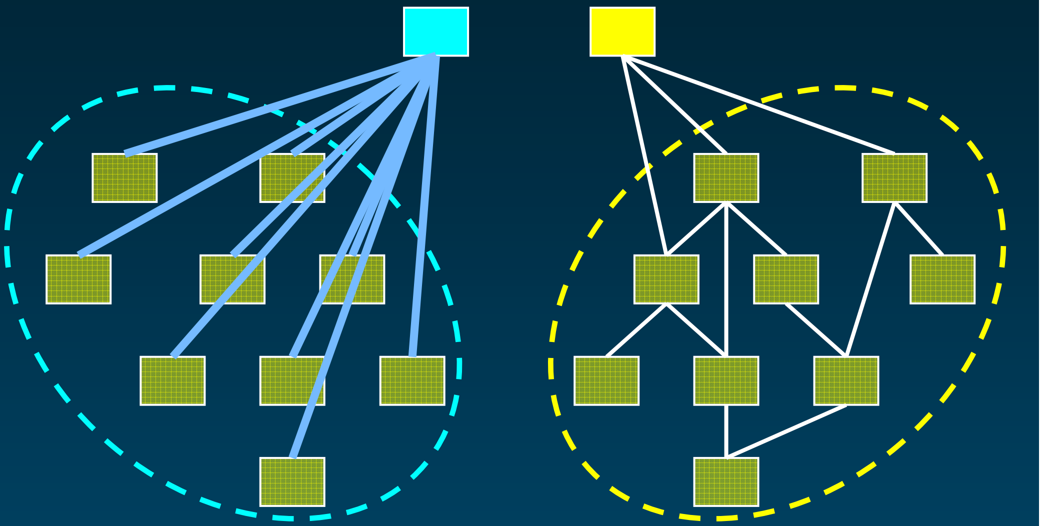
- Map semantic type names to the Metathesaurus
  - Exact match
  - After normalization if necessary
- Adapt semantic type (ST) names
  - Decompose coordinated ST names
    - *Injury or Poisoning* → *Injury* + *Poisoning*
  - Distribute modifiers as required
    - *Body Space or Junction* → *Body Space* + *Body Junction*

Lexical similarity Results

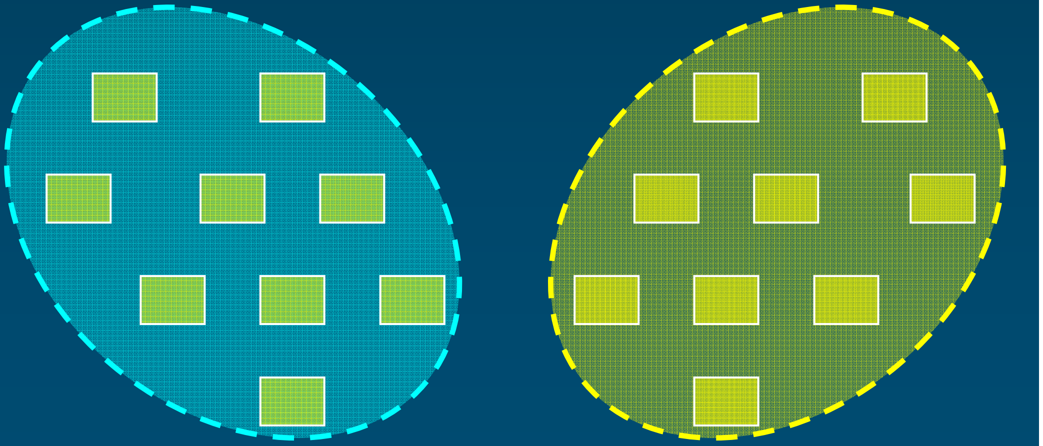
- 135 semantic types
  - 32 coordinated with *or*
- 172 names after decomposition
- Mapping to UMLS concepts and manual review
  - 106 unique and relevant
  - 10 multiple (requiring disambiguation)
  - 66 names failed to be mapped (e.g., *Biologic Function*, *Temporal Concept*)

❷ Conceptual similarity Method

- Semantic type
  - List of all concepts having this semantic type
- Concept
  - List of all descendants



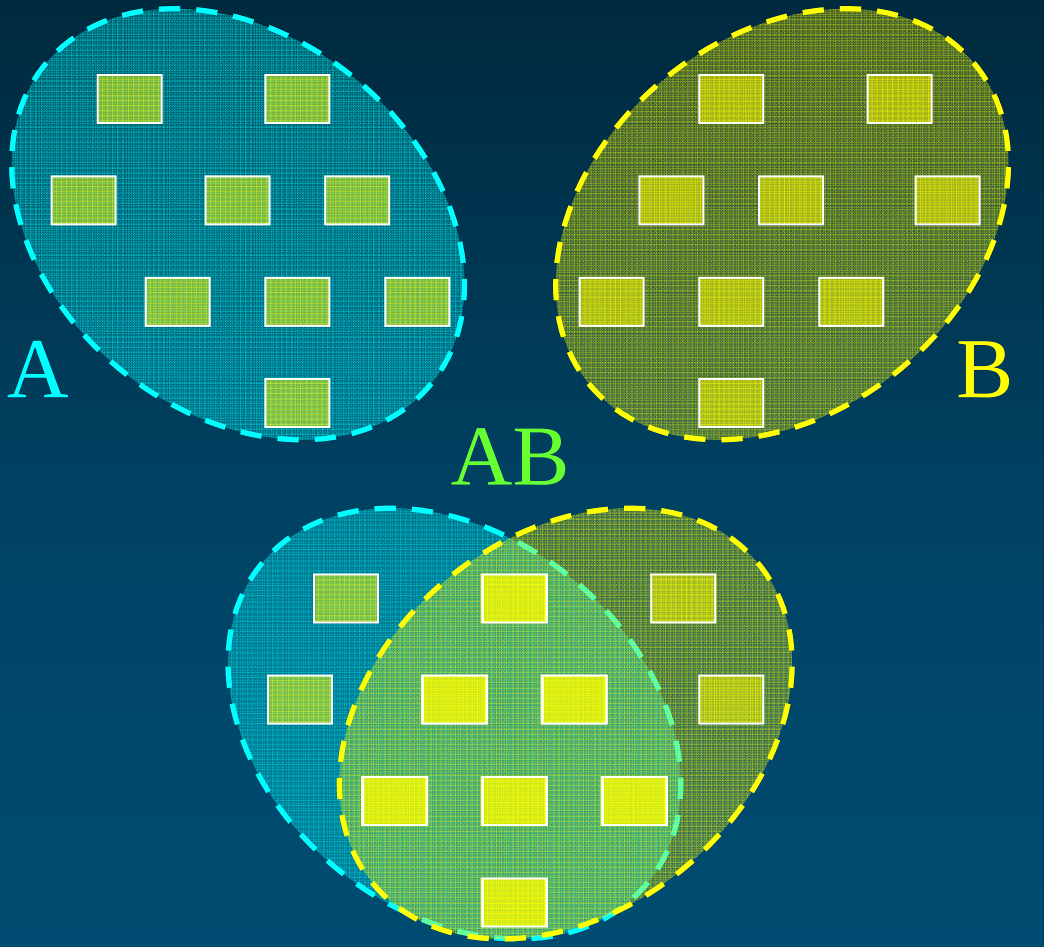
- Comparing the 2 sets
  - Intersection of the 2 sets
  - Similarity measures
    - Cosine
    - Jaccard
    - Dice



Cosine similarity measure Method

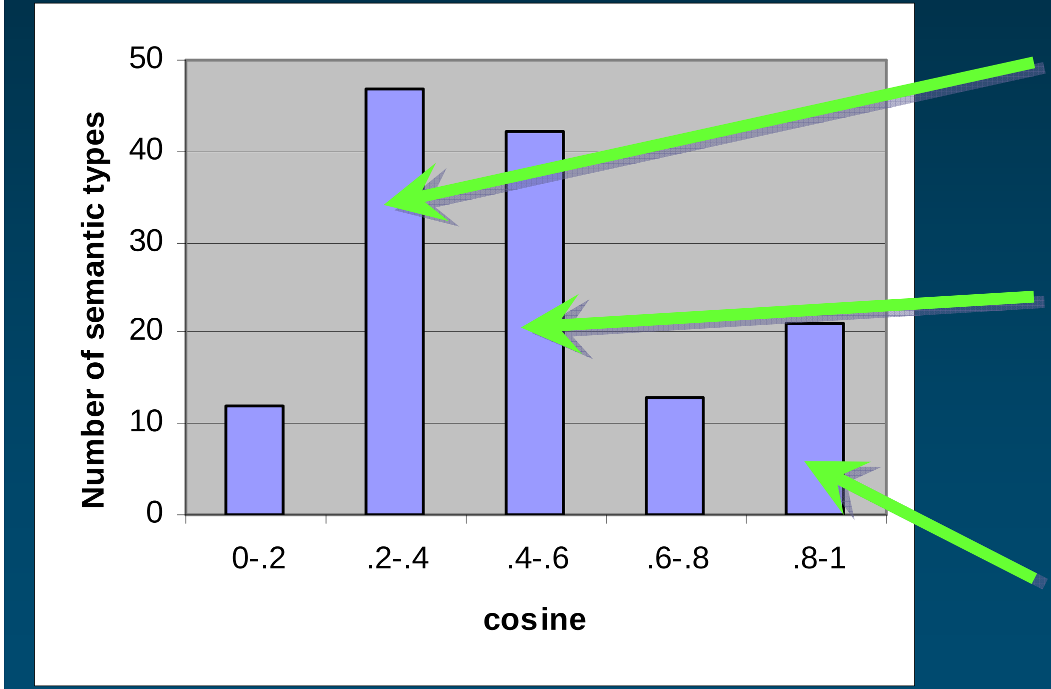
$$Sim_{\cos} = \frac{AB}{\sqrt{A*B}}$$

$$Sim_{\cos} = \frac{7}{\sqrt{9*9}} = .78$$



Conceptual similarity Results

- Top cosine values for each semantic type ranged from .0094 to .9943



Sim (*Immunologic Factor*, *Immunology*) = .3242

Sim (*Gene or Genome*, *Cancer genes*) = .6781

Sim (*Gene or Genome*, *Genes*) = .6466

Sim (*Reptile*, *Lepidosauria*) = .9729

Sim (*Amphibian*, *Amphibia*) = .9943

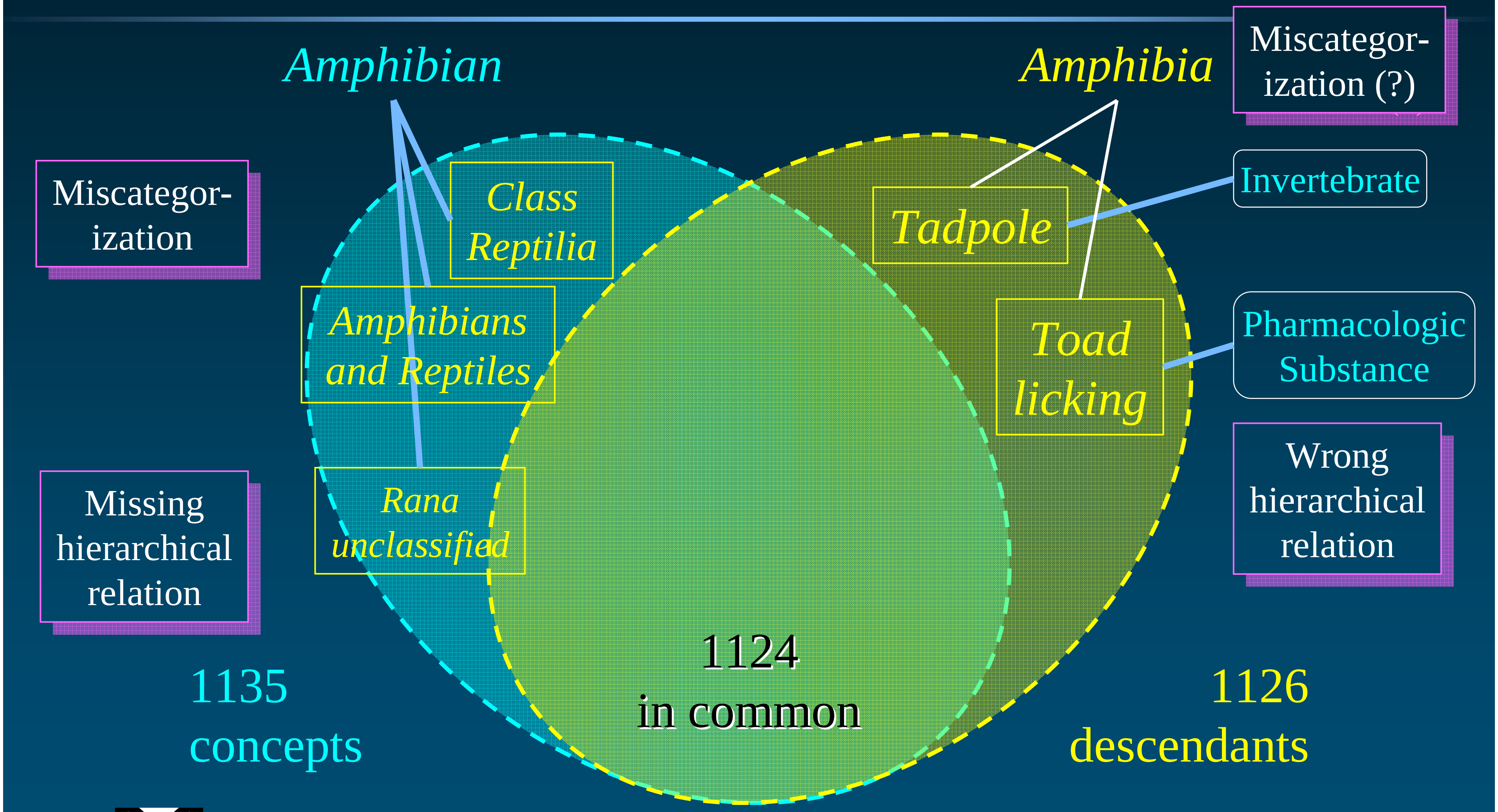
Lexical vs. conceptual similarity

- 106 relevant mappings obtained by lexical similarity between a semantic type name and a Metathesaurus concept
  - In 60 cases, the concept mapped to lexically was among the top 25 candidates identified by conceptual similarity
  - 10 concepts mapped to lexically had no descendants
  - In 36 cases, lexical similarity with limited conceptual similarity

Applications

- Auditing consistency
  - Hierarchical relations and the categorization of concepts are expected to be consistent
- Extending the semantic network downwards
  - Using the descendants of the corresponding high-level concepts as candidates

Auditing consistency



Amphibian

Class Reptilia

Amphibians and Reptiles

Rana unclassified

Amphibia

Tadpole

Toad licking

Miscategorization

Missing hierarchical relation

Miscategorization (?)

Invertebrate

Pharmacologic Substance

Wrong hierarchical relation

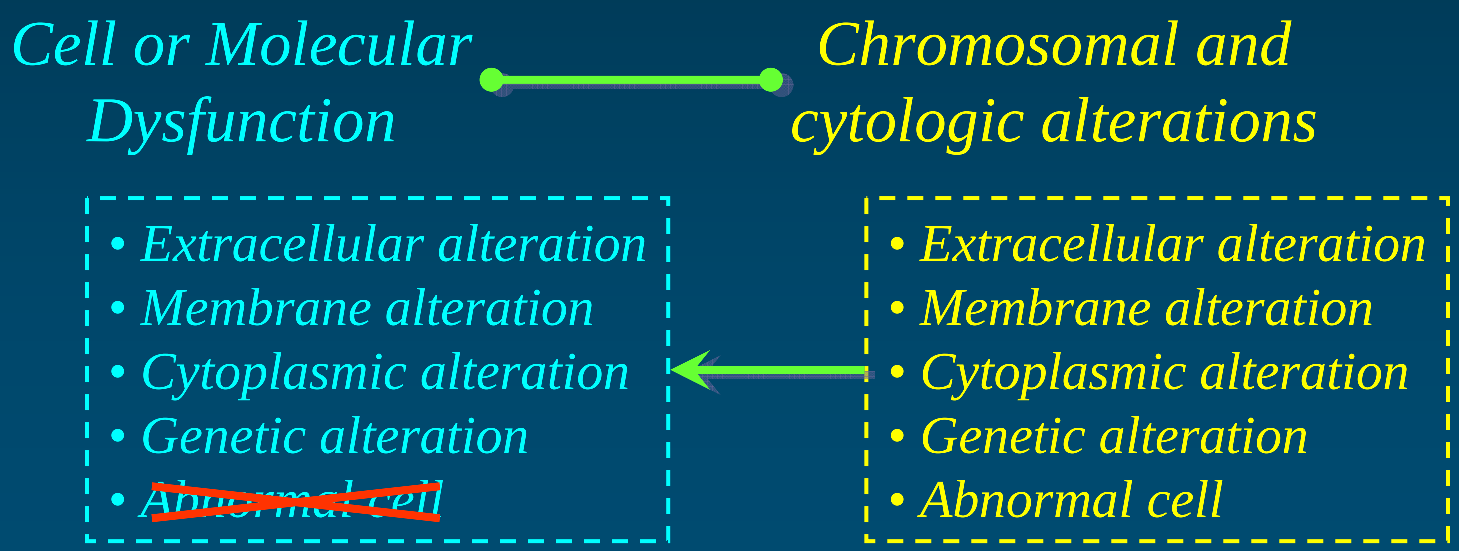
1135 concepts

1124 in common

1126 descendants

Extending the semantic network

- Select the concept corresponding to a given semantic type (ST)
- The first-generation descendants of this concept become candidate children for the ST



Cell or Molecular Dysfunction

Chromosomal and cytologic alterations

- *Extracellular alteration*
- *Membrane alteration*
- *Cytoplasmic alteration*
- *Genetic alteration*
- ~~*Abnormal cell*~~

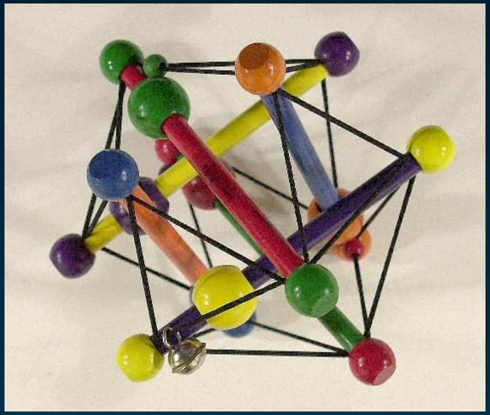
- *Extracellular alteration*
- *Membrane alteration*
- *Cytoplasmic alteration*
- *Genetic alteration*
- *Abnormal cell*

Limitations

- Lexical similarity
  - False positives (polysemy)
  - False negatives (missing synonyms)
- Conceptual similarity
  - Difficult to set a threshold
- Applications
  - Require some degree of manual intervention

Conclusions

- Aligning two UMLS knowledge sources
  - Metathesaurus
  - Semantic Network
- Two complementary approaches
  - Lexical similarity
  - Conceptual similarity
- Application to
  - Auditing consistency
  - Extending the semantic network downwards



Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

NATIONAL LIBRARY OF MEDICINE

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA